Self-supervision for data-driven anomaly detection at the LHC

EuCAIFCon 2024 - Amsterdam

Luigi Favaro

In collaboration with B. Dillon, F. Feiden, M. Krämer, T. Modak, T. Plehn, J. Rüschkamp

May 2, 2024

UNIVERSITÄT HEIDELBERG
ZUKUNFT
SEIT 1386

Funded by
DFG Deutsche Forschungsgemeinschaft
German Research Foundation

Table of Contents

Anomaly searches

1 Introduction

- LHC does and will generate large amount of data;
- We are still looking for BSM physics;
- No clear anomalies in the near future . . .
  $\longrightarrow$ keep exploring with direct searches is not feasible;

Anomaly searches

1 Introduction

- LHC does and will generate large amount of data;
- We are still looking for BSM physics;
- No clear anomalies in the near future . . .

  $\longrightarrow$ keep exploring with direct searches is not feasible;

model agnostic searches

Anomaly searches

1 Introduction

- LHC does and will generate large amount of data;
- We are still looking for BSM physics;
- No clear anomalies in the near future . . .
  $\longrightarrow$ keep exploring with direct searches is not feasible;

model agnostic searches **no loss in sensitivity**

Anomaly searches

1 Introduction

- LHC does and will generate large amount of data;
- We are still looking for BSM physics;
- No clear anomalies in the near future . . .

  $\longrightarrow$ keep exploring with direct searches is not feasible;

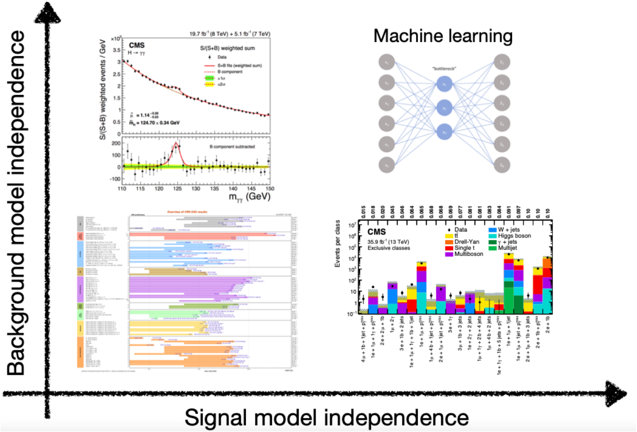
model agnostic searches

no loss in sensitivity

Have we fully explored the collected data?

Anomaly searches

1 Introduction



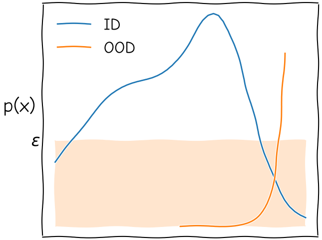
credits to M. Krämer, cf. Karagiorgi, Kasieczka, Kravitz, Nachman, Shih, arXiv:2112.03769 [hep-ph]

Density estimation

1 Introduction

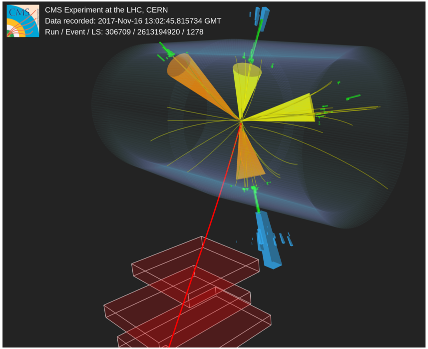
- Density estimates used as agnostic anomaly scores;
- Only focus on density estimation of background:
  - can be used directly on data;
  - sensitivity requires dealing with large input spaces;
- Improved representation of the data is key. . .

$\longrightarrow$ Physics motivated preprocessing/observables.



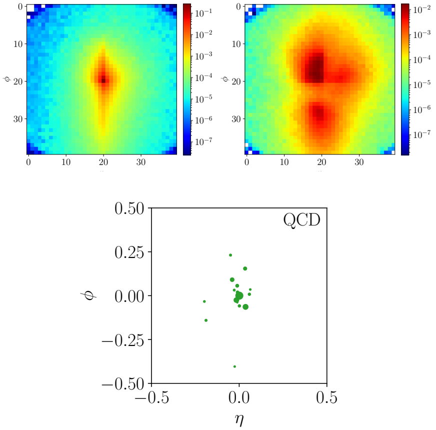
$$OOD = \{x | p(x) < \epsilon\}$$

Datasets
1 Introduction



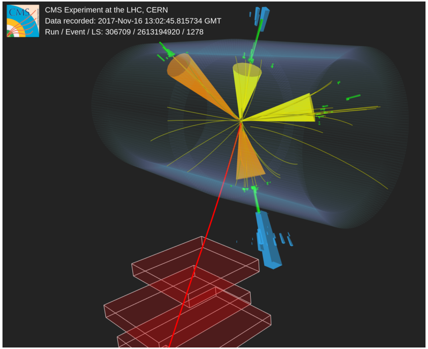
*from CMS Open Data

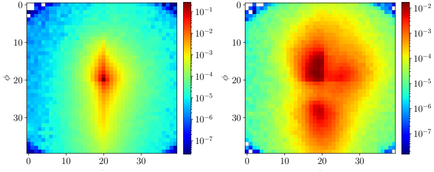
Reco events

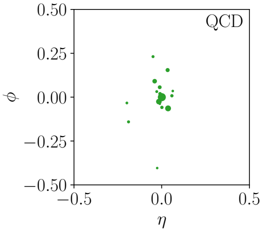


Jet substructure

Datasets
1 Introduction







*from CMS Open Data

Set of objects of variable length with features (p_T, η, ϕ)

Table of Contents

Self-supervision

2 Contrastive Learning

*also discussed in Patrick's talk here

- Preprocessing solves many problems with neural networks;
  - NNs are not invariant to physical symmetries in data;
  - NNs like numbers of $O(1)$;
- Self-supervision: during training we use pseudo-labels, not truth labels;

$\longrightarrow$ task used to create new representations/observables;

Self-supervision

2 Contrastive Learning

*also discussed in Patrick's talk here

- Preprocessing solves many problems with neural networks;
  - NNs are not invariant to physical symmetries in data;
  - NNs like numbers of $O(1)$;
- Self-supervision: during training we use pseudo-labels, not truth labels;

  $\longrightarrow$ task used to create new representations/observables;

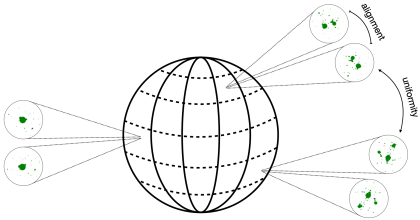
Key aspects of representations:

- Invariance to certain transformations of the jet/event
- Discriminative power

In CLR we construct a mapping to a new representation space

*as in JetCLR, arXiv:210804253

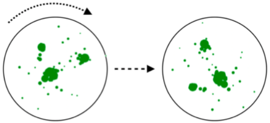
- Pseudo-labels are defined from pairs:
  - $\{(x_i, x'_i)\}$: positive pair
    $\rightarrow$ alignment/invariance;
  - $\{(x_i, x_j) \cup (x_i, x'_j))\}$: negative pair
    $\rightarrow$ uniformity/discriminative;
- $f : \mathcal{R} \rightarrow \mathcal{Z}$ is a transformer-encoder;
- $s(\cdot, \cdot)$ cosine distance in rep. space.



$$\mathcal{L} = -\log \frac{exp(s(z_i, z'_i)/\tau)}{\sum \mathcal{I}_{i \neq j}[exp(s(z_i, z_j)/\tau) + exp(s(z_i, z'_j)/\tau)]}$$

Augmentation: any transformation (e.g. rotation) of the original jet

- Example:
  $\longrightarrow$ detector invariance under rotations;



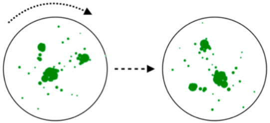
CLR for anomaly detection

2 Contrastive Learning

Augmentation: any transformation (e.g. rotation) of the original jet

- Example:
  $\longrightarrow$ detector invariance under rotations;

- Background data does not known BSM features.

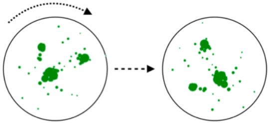


Can we train a transformer-encoder network only on background events?

- Possible, but with no guarantee to learn representations sensitive to new physics;

Augmentation: any transformation (e.g. rotation) of the original jet

- Example:
  $\longrightarrow$ detector invariance under rotations;



- Background data does not known BSM features.

Can we train a transformer-encoder network only on background events?

- Possible, but with no guarantee to learn representations sensitive to new physics;

Introduce general BSM motivated anomalous representations z^*

*from "Anomalies, representations, and self-supervision", arXiv:2301.04660

Contrastive Learning for anomaly detection:

- anomalous pairs: $\{(x_i, x_i^*)\}$ where x_i^* comes from a different set of augmentations;
- These are motivated by BSM features, general, and signal agnostic.

$$\mathcal{L}_{aCLR} = -\log \frac{exp(s(z_i, z_i') - s(z_i, z_i^*)/\tau)}{\sum \mathcal{I}_{i\neq j}[exp(s(z_i, z_j)/\tau) + exp(s(z_i, z_j')/\tau)]}$$

$$\mathcal{L}_{aCLR+} = -\log \exp^{(s(z_i, z_i') - s(z_i, z_i*))/\tau} = \frac{s(z_i, z_i^*) - s(z_i, z_i)}{\tau}$$

Table of Contents

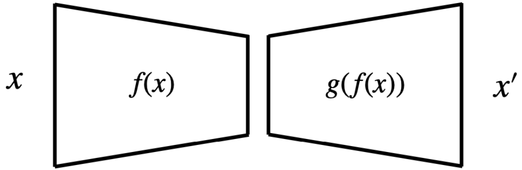
Anomaly scores

3 Event anomalies

- Cut-based OOD downstream task as a benchmark;
- (Normalized)AutoEncoder based anomaly score:

$$E_\theta = MSE(x, D(E(x))) \qquad p_\theta(x) = \frac{e^{-E_\theta(x)}}{\Omega};$$

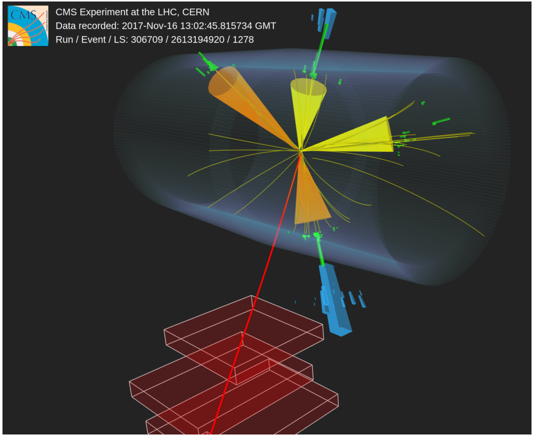
- for more details see arXiv:2206.14225;
- or posters IDs 146,95 for applications within CMS.



The corresponding anomaly score will be (approx) invariant to the augmentations

Reco events

3 Event anomalies



Reco events

3 Event anomalies

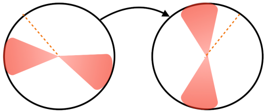
- Background data:
  - $W \to l\nu$ (59.2%);
  - $Z \to ll$ (6.7%);
  - QCD multijets (33.8%);
  - $t\bar{t}$ production (0.3%);
- Benchmark signals:
  - $A \to 4l$;
  - $h^0 \to \tau\tau$;
  - $h^+ \to \tau\nu$;
  - $LQ \to b\nu$;

- Event-level reconstructed objects:
  - (p_T, ϕ) of missing transverse energy;
  - (p_T, η, ϕ) of e, μ, and jets;
- Selection cuts:
  - one lepton with $p_T > 23$ GeV;
  - $|\phi| \in [-\pi, \pi]$, rescaled in $[-1, 1]$;
  - $|\eta| < 3, 2.1, 4$ for e, μ, jets, rescaled in $(-1, 1)$;
- empty entries are zero-padded.

Augmentations

3 Event anomalies

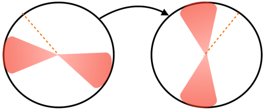
- Azimuthal rotations;
- (η, ϕ) and energy smearing.

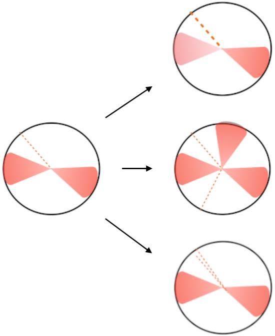


Augmentations

3 Event anomalies

- Azimuthal rotations;
- (η, ϕ) and energy smearing.

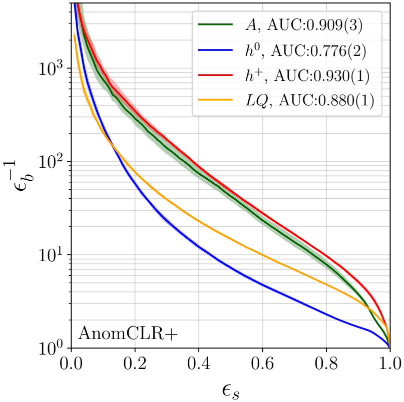


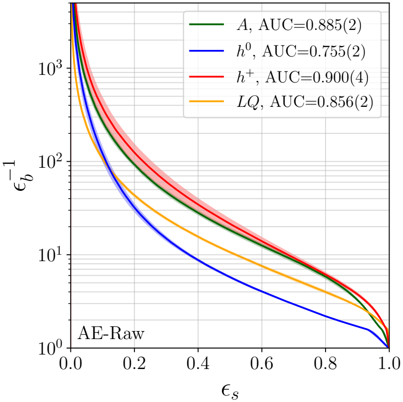


- p_T and MET shifts;
- random multiplicity shifts;
- multiplicity shifts with MET and p_T constant.

Effect of new representation

3 Event anomalies



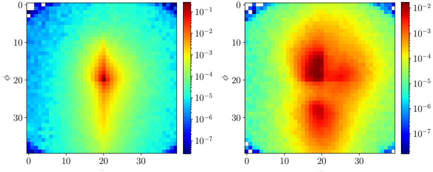


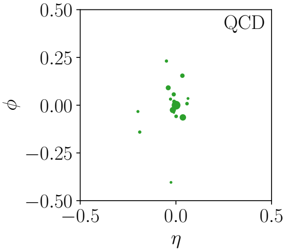
Better performance on all the BSM signals

Table of Contents

Jet substructure

4 Representing dark showers





Dark showers

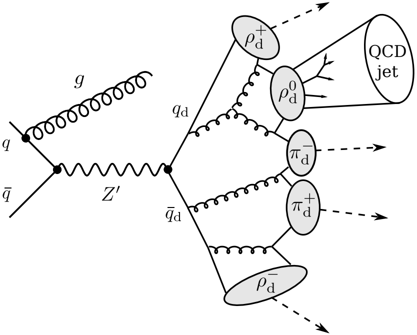
*from "Semi-visible jets, energy-based models, and self-supervision", arXiv:2312.03067

New physics hidden in jet substructure;

Benchmark signal: semi-visible jets

- $Z' = 2\text{TeV}$ dark sects mediator;
- q_d dark quarks charged under $SU(3)_d$;
- $m_{q_d} = 500\text{MeV}$;
- $\Lambda = m_{\pi_d} = m_{\rho_d} = 5\text{GeV}$;

QCD-like showers with fraction of invisible particles



*studied in arXiv:2006.08639

Dark showers

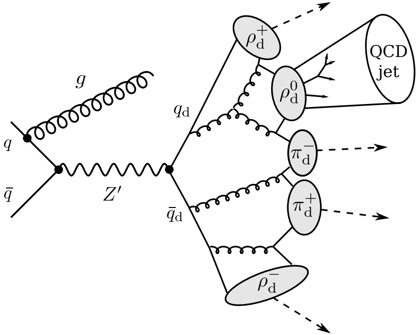
4 Representing dark showers

*from "Semi-visible jets, energy-based models, and self-supervision", arXiv:2312.03067

New physics hidden in jet substructure;

Benchmark signal: semi-visible jets

- Jet constituents:
  - (p_T, η, ϕ) of each constituent;
  - $p_T \in [150, 350]$ GeV, $|\eta_j| < 2$;
- anti-kt clustering $\Delta R = 0.8$;
- empty entries are zero-padded.

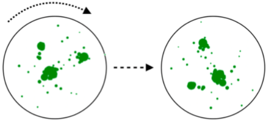


$r_{\text{inv}} = 0.75$, $m_\mathrm{d} = 5\text{GeV} \longrightarrow$ referred to as "Aachen" model.

*studied in arXiv:2006.08639

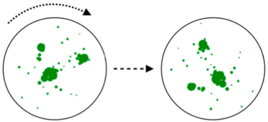
Augmentations

rotations in $[0, 2\pi]$:

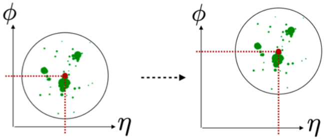


Augmentations

rotations in $[0, 2\pi]$:

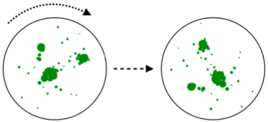


translations in $[\eta, \phi]$:

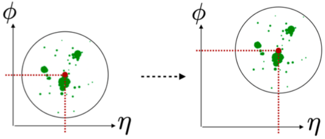


Augmentations

rotations in $[0, 2\pi]$:



translations in $[\eta, \phi]$:



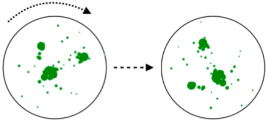
permutation invariance:

$$f(x) = f(\mathcal{S}_n(x))$$

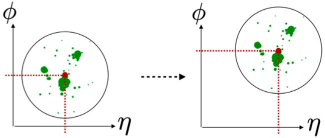
Augmentations

4 Representing dark showers

rotations in $[0, 2\pi]$:



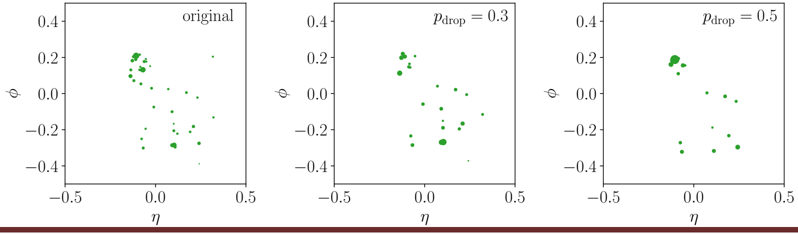
translations in $[\eta, \phi]$:



permutation invariance:

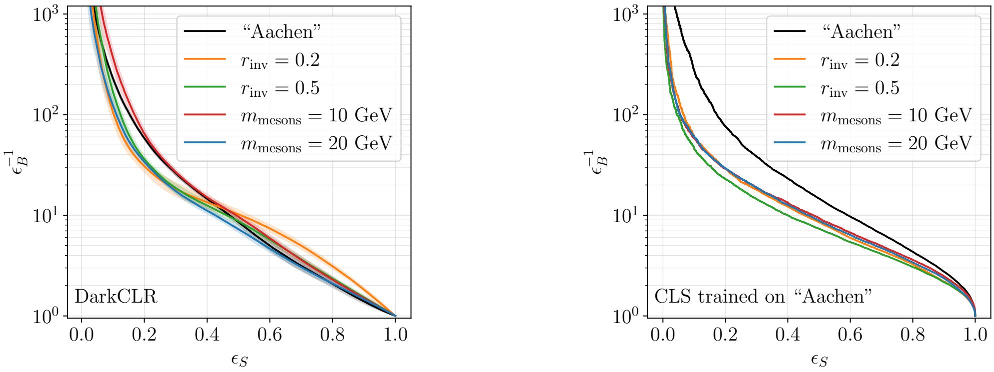
$$f(x) = f(\mathcal{S}_n(x))$$

Applying p_{drop} to a QCD jet:



Robustness of darkCLR

4 Representing dark showers



Representations generalize over different pheno parameters

- ML can lead the future of anomaly searches;

Conclusions

- ML can lead the future of anomaly searches;
- We can use CLR to learn powerful representations:
  - approximately invariant under transformations;
  - sensitive to BSM effects;

EUROPEAN AI FOR FUNDAMENTAL PHYSICS CONFERENCE EuCAIFCon 2024

Conclusions

5 Conclusions

- ML can lead the future of anomaly searches;
- We can use CLR to learn powerful representations:
  - approximately invariant under transformations;
  - sensitive to BSM effects;
- Two examples of anomaly detection with CLR:
  - Event-level reconstructed objects;
  - Semivisible jets detection;

EUROPEAN AI FOR FUNDAMENTAL PHYSICS CONFERENCE EuCAIFCon 2024

Conclusions

5 Conclusions

- ML can lead the future of anomaly searches;
- We can use CLR to learn powerful representations:
  - approximately invariant under transformations;
  - sensitive to BSM effects;
- Two examples of anomaly detection with CLR:
  - Event-level reconstructed objects;
  - Semivisible jets detection;

Outlook:

- Have a more interpretable latent space;
- More detailed study of systematic uncertainties.

EUROPEAN AI FOR FUNDAMENTAL PHYSICS CONFERENCE EuCAIFCon 2024

Self-supervision for data-driven anomaly detection at the LHC

EuCAIFCon 2024 - Amsterdam

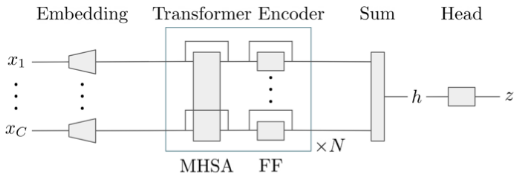
Thank you for listening!

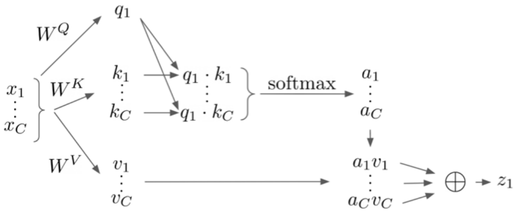
Any questions?

Backup

Transformer Encoder

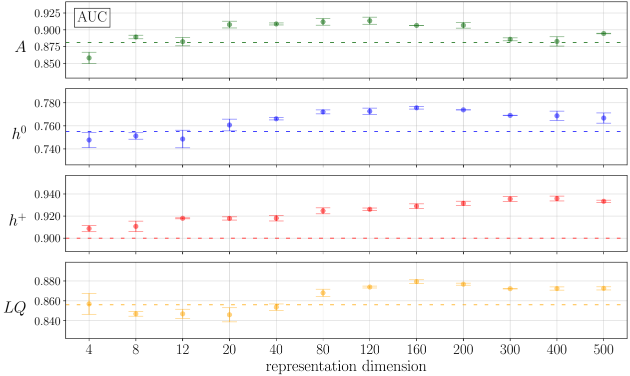
6 Backup

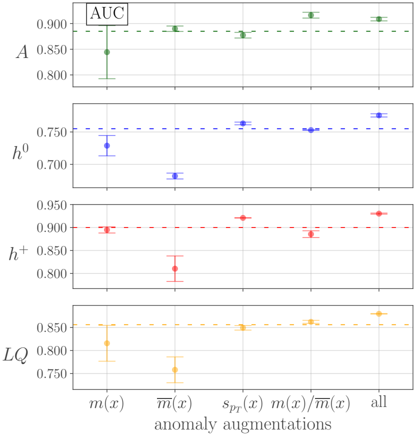




Ablation studies

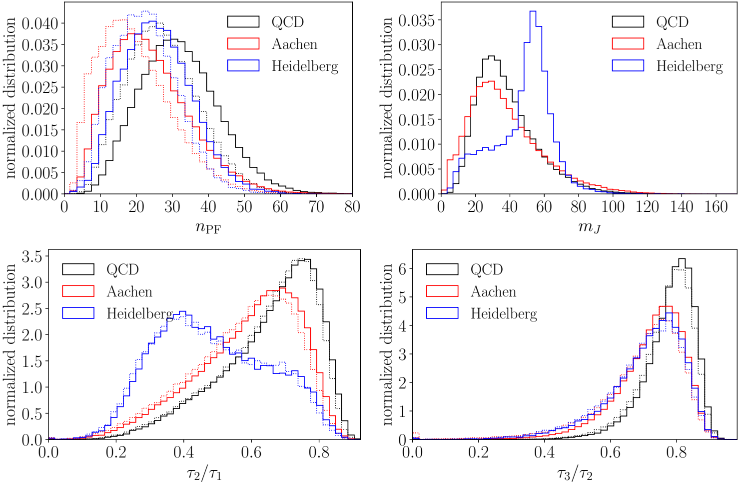
6 Backup

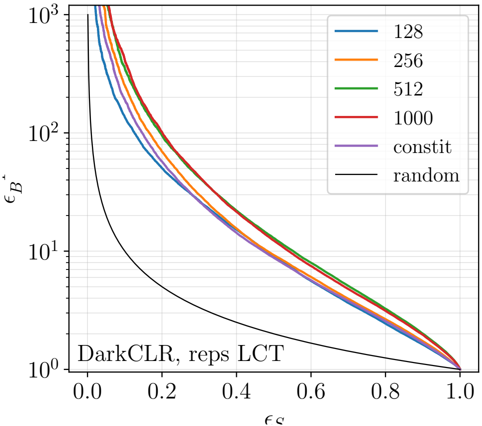




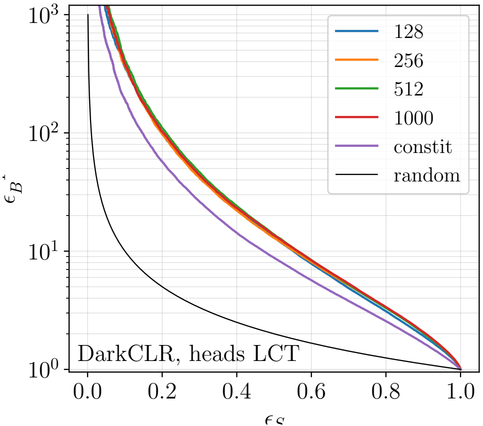
High-level features

6 Backup





10^3
10^2
10^1
10^0
ϵ_B^{-1}
0.0
0.2
0.4
0.6
0.8
1.0
ϵ_S
128
256
512
1000
constit
random
DarkCLR, reps LCT



10^3
10^2
10^1
10^0
ϵ_B^{-1}
0.0
0.2
0.4
0.6
0.8
1.0
ϵ_S
128
256
512
1000
constit
random
DarkCLR, heads LCT